UCL Energy Institute

Characterising Uncertain Long-term Decarbonisation Pathways with Clustering Algorithms

Dr. Pei-Hao Li (p.li@ucl.ac.uk)

Mr. Steve Pye

Dr. Ilkka Keppo

UCL Energy Institute

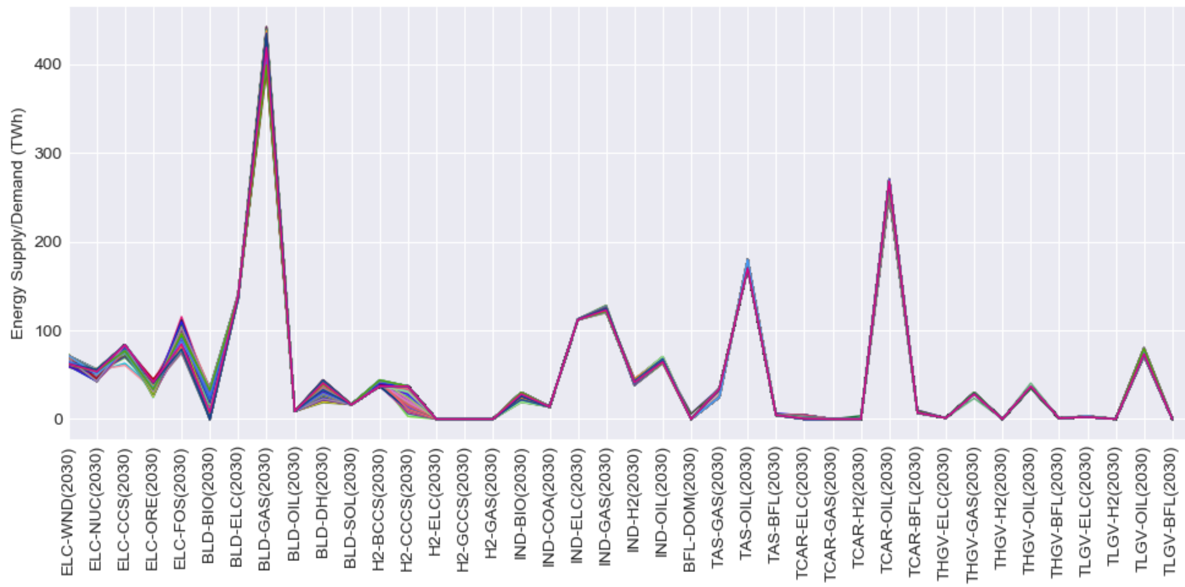
2019 IEW, Paris, 4th June, 2019

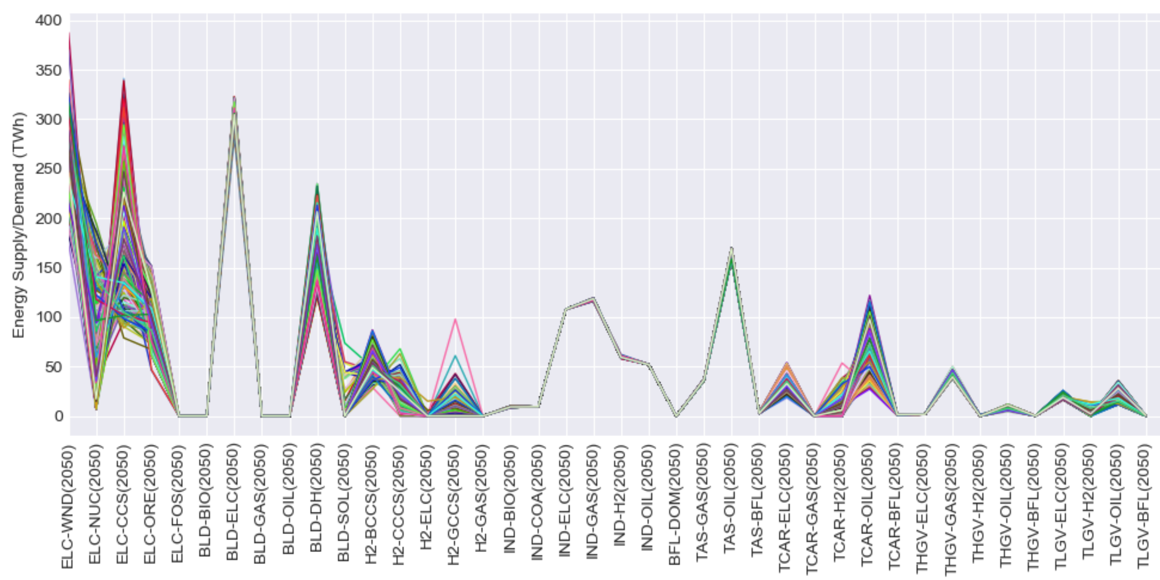
Outline

- Motivation
- Research procedure
- Uncertain long-term decarbonisation pathways
  - Energy System Modelling Environment (ESME)
- Clustering algorithms
  - (K-means, hierarchical clustering, Gaussian mixture model, spectral clustering, density-based clustering)
- Proximity matrix and transformation
- Performance evaluation of clustering algorithms
- Representative long-term pathways
- Coevolution of metrics
- Conclusions

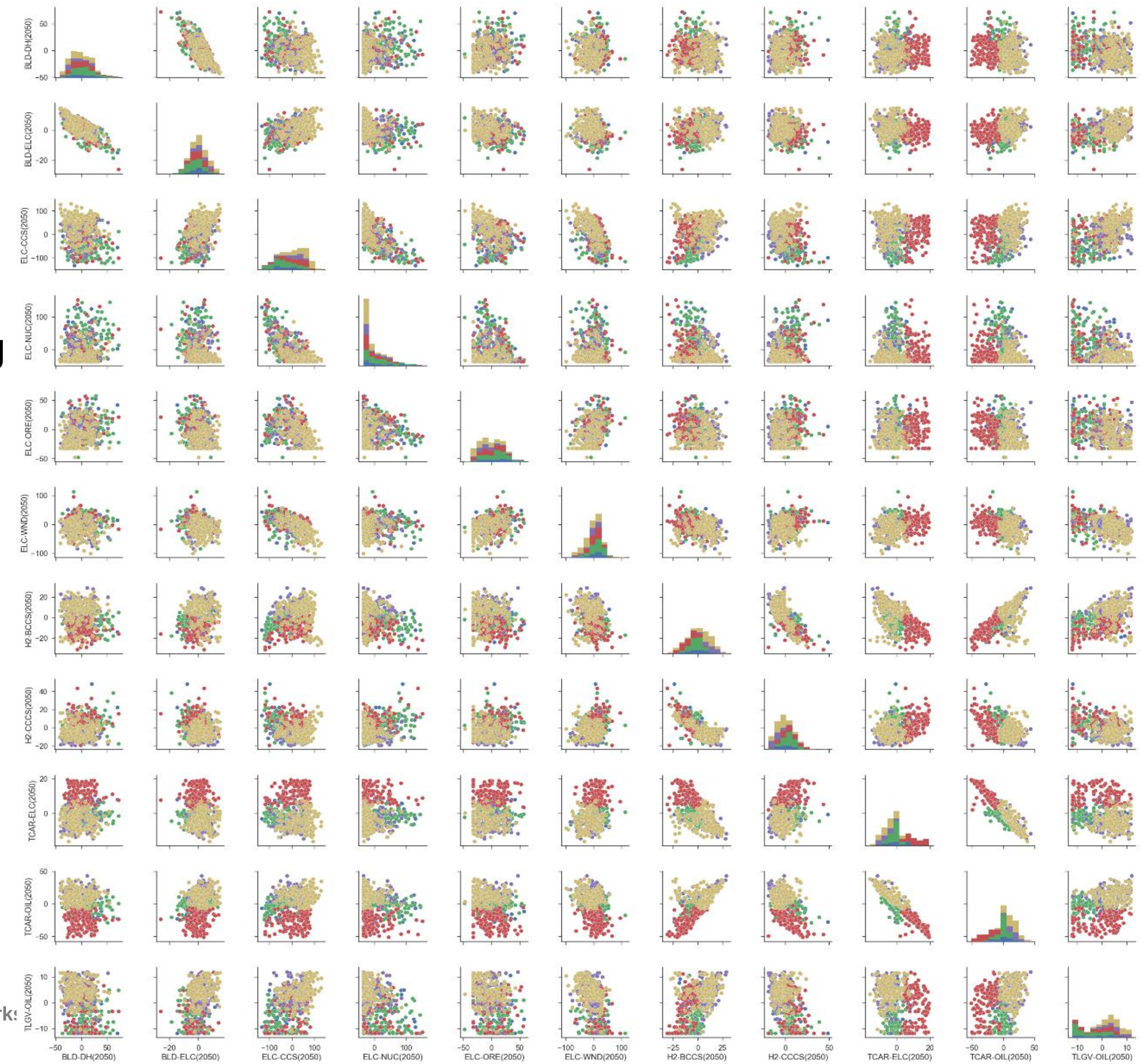
What if we have a huge number of decarbonisation pathways?

Distribution of metrics in 2030 and 2050 across pathways





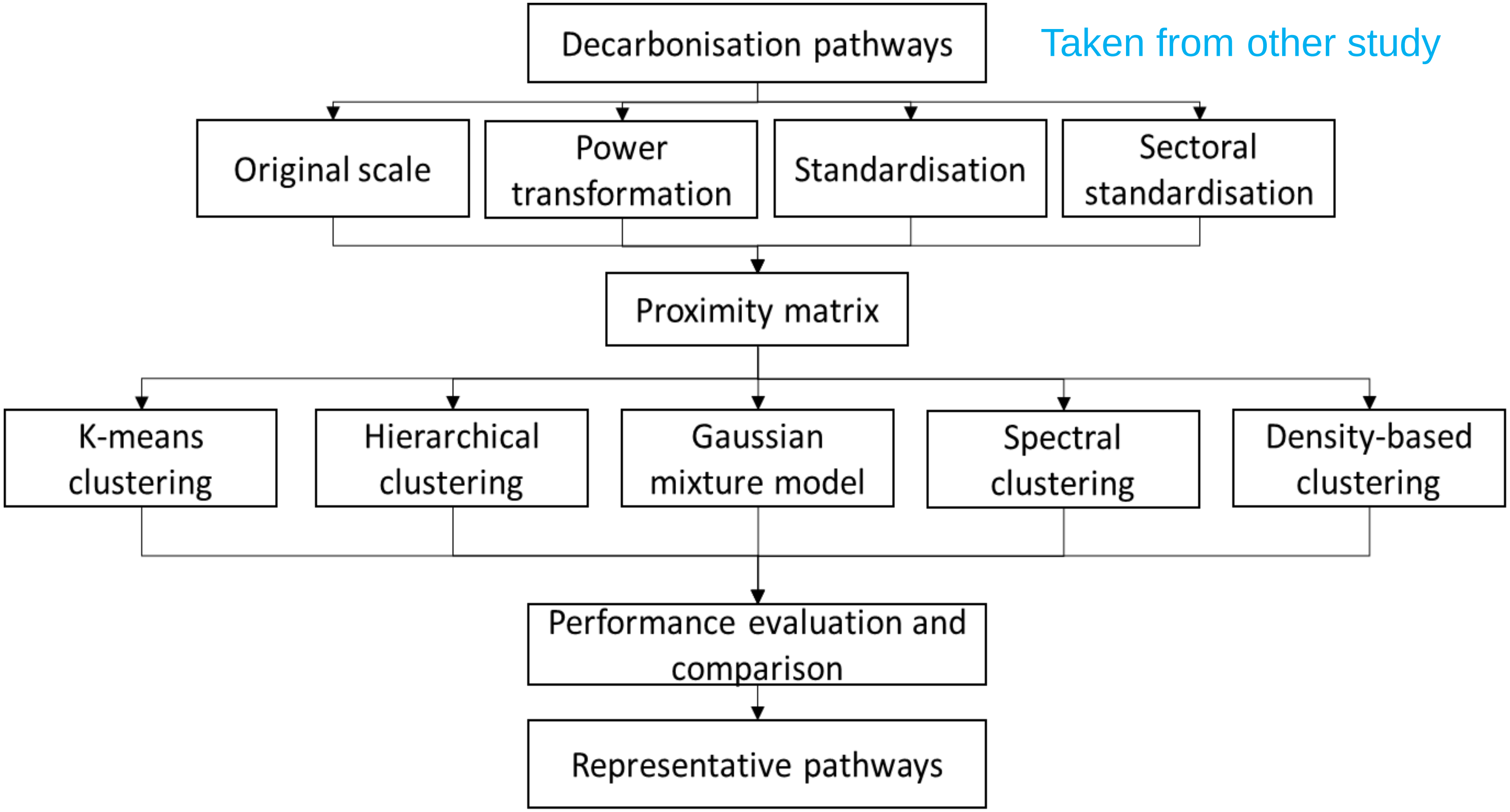
Scatter plots showing relationship between key metrics in 2050



Motivation

- Whole energy system models, such as UK TIMES (~2000 techs), Irish TIMES (1700 techs), are essential tools to help policy-makers decide long-term decarbonisation pathways
- However, future is highly uncertain! Model is extremely sensitive to input assumptions!
  - Technology cost, resource availability, etc.
- **Uncertainty analysis**
  - Global sensitivity analysis (Fais et al., 2016; Pye et al., 2015)
  - Modelling to generate alternatives (Price and Keppo, 2017; Li and Trutnevyte, 2017; DeCarolis et al., 2016)
- Unmanageable number of pathways!
- How can we find patterns out of these to support policy-making?
  - Representative pathways
    - Represent the whole set of pathways
    - Different enough from each other
  - Relationship between technologies
- ***Challenges: extremely high-dimensional and unlabeled!!***
- **Clustering algorithms** come to help!

Research Procedure



ESME
Energy System Modelling Environment

Energy System Modelling Environment

- Systems optimisation via linear programming

>300 technologies

Probabilistic treatment of key uncertainties

TECHNOLOGY ROADMAPS

DEMAND FOR ENERGY

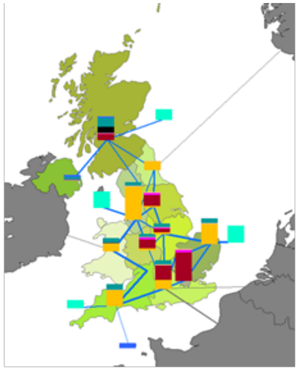
ENERGY RESOURCES

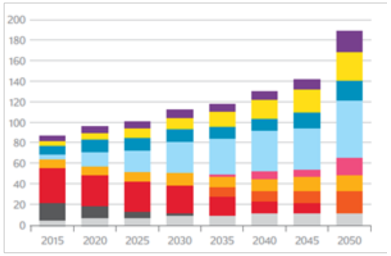
Least cost optimisation

Pathway and supply chain constraints to 2050

ENERGY SYSTEM BLUEPRINTS

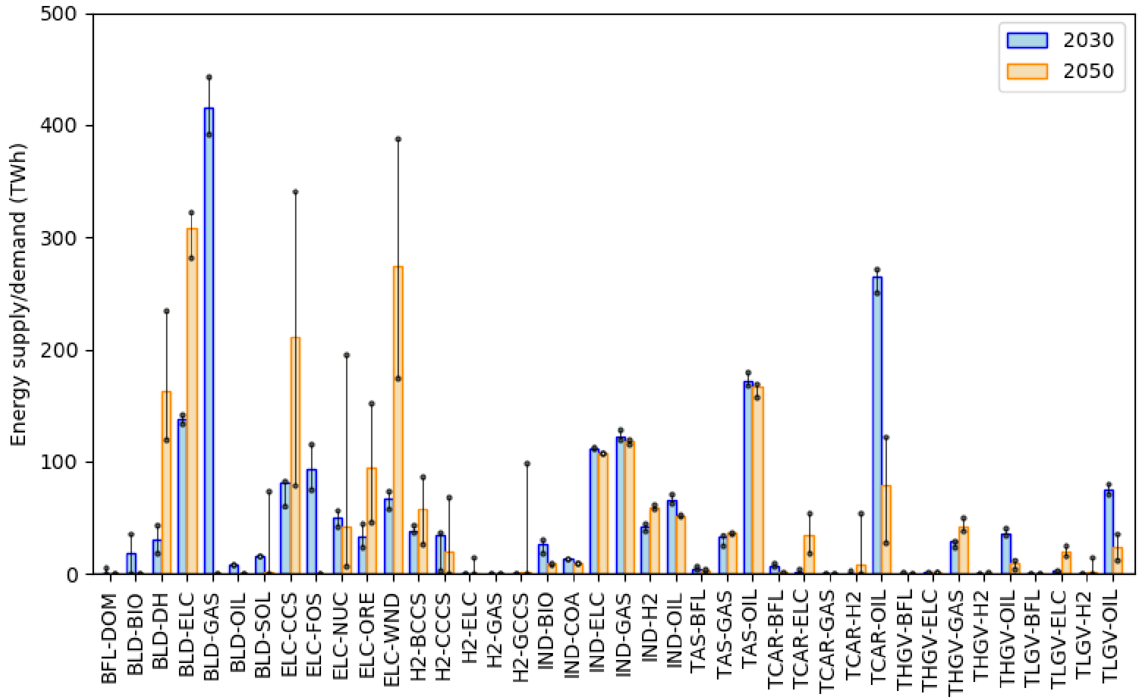
Spatially explicit UK regions





Uncertain Long-term Decarbonisation Pathways

- Taken from existing study (Pye et al., 2019)
- -80% GHG reduction in 2050 (rel. to 1990), -53% in 2030
- Uncertain characteristics
  - Capital costs
  - Commodity costs
  - Build rates
  - Resource availability
- Variation of parameters (by 2050)
  - Mature: +/-10%
  - New: +/-30%
  - Novel/emerging: +/-50%
- Monte Carlo technique
- 600 pathways

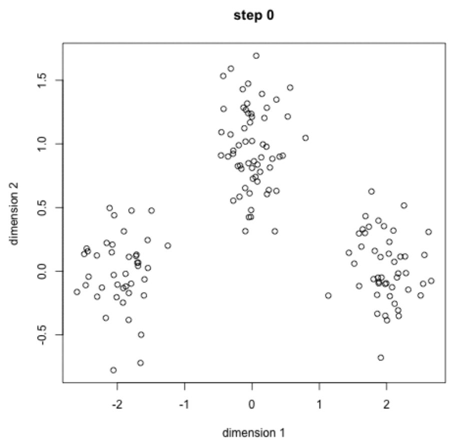


Pye, S, Li, P-H, Keppo, I, O'Gallachoir, B. 2019. Technology interdependency in the United Kingdom's low carbon energy transition, Energy Strategy Reviews, 24, 314-330.

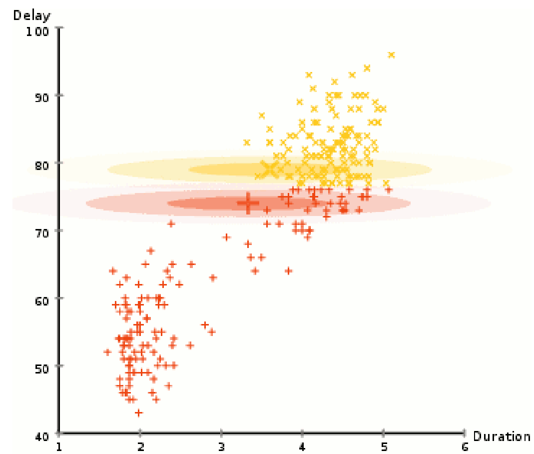
Clustering algorithms in action

Need predefined number of clusters

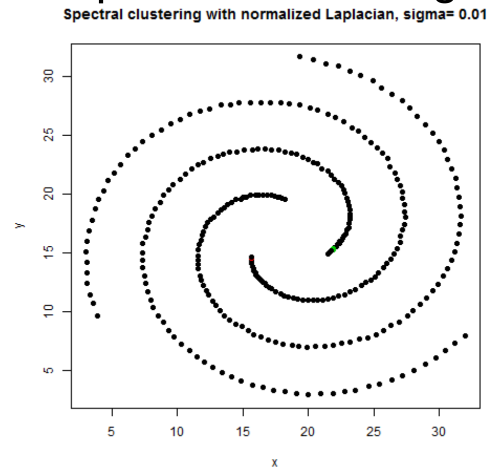
K-means



GMM

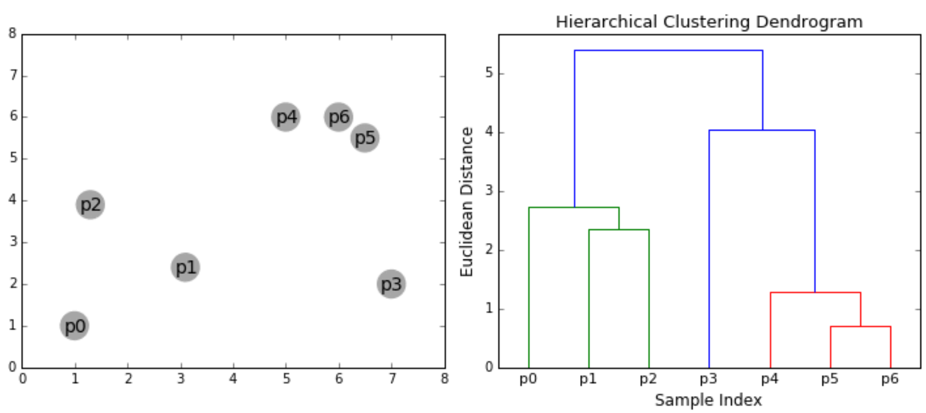


Spectral clustering

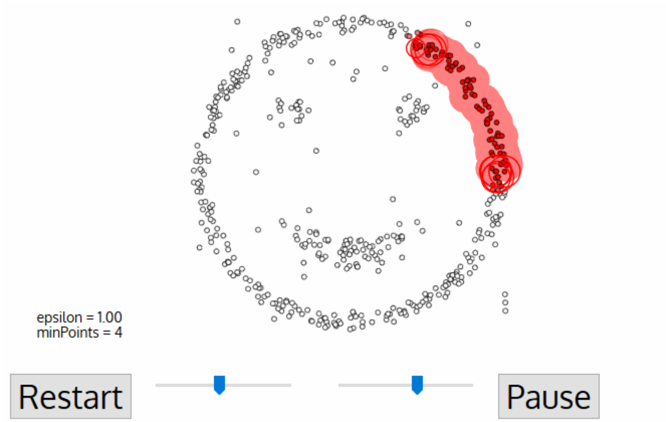


Don't need predefined number of clusters

Hierarchical clustering



DBSCAN



Source: George Seif, The 5 Clustering Algorithms Data Scientists Need to Know. https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68
https://rstudio-pubs-static.s3.amazonaws.com/199446_36b75293d1544636a278369aefd9a094.html

Proximity Matrix

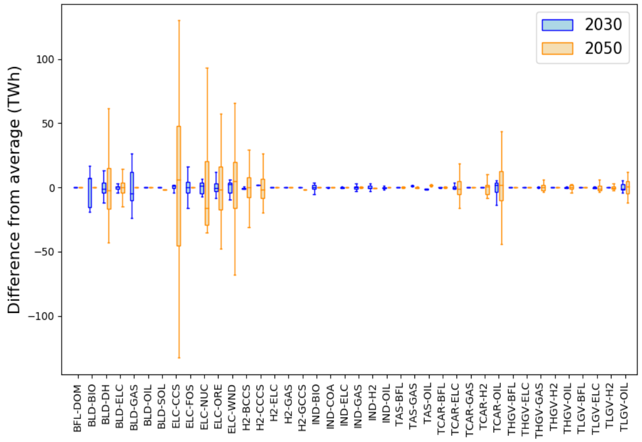
- **Proximity matrix:** representing the similarity between pathways for clustering analysis
- **Proximity:** summation of differences of metrics between pathways over modelling years
  - $d_{i,j} = \sqrt{\sum_y \sum_{m=1}^{M} (x_{i,y,m} - x_{j,y,m})^2}$
  - Only consider 2030 and 2050 for simplification

- Issues
  - Pathway might be dominated by a few key metrics, such as nuclear
  - Need pathway sets with various distribution characterstics for robustness testing
- Metric transformation is thus applied!

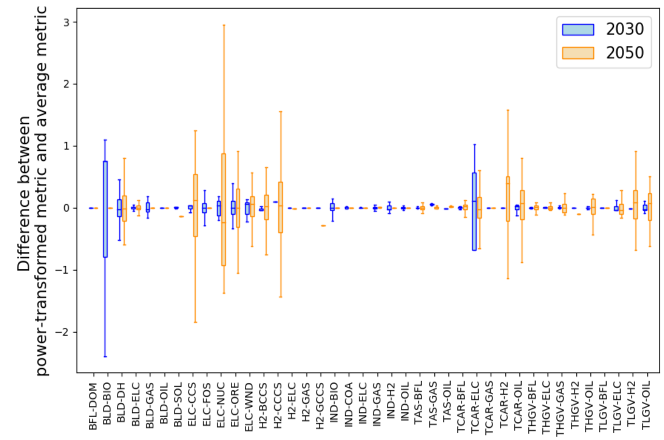
Metric Transformation

Difference between transformed metric and average metrics

No transformation

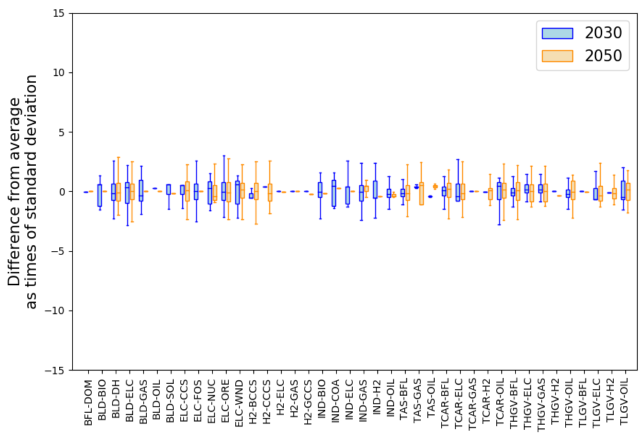


Power



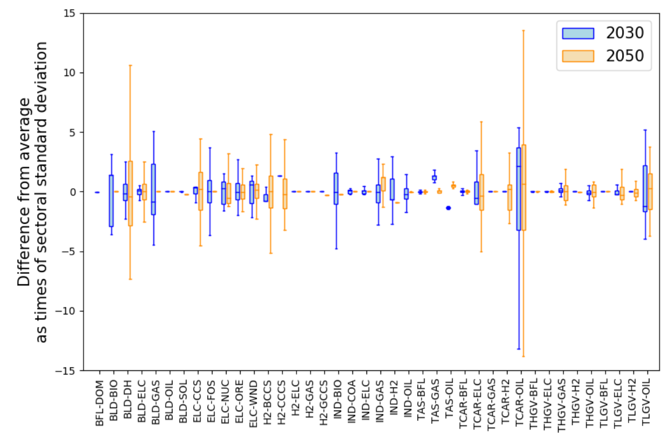
Reduce the influence of metrics with extreme high values

Standarisation



Treat every metric equally

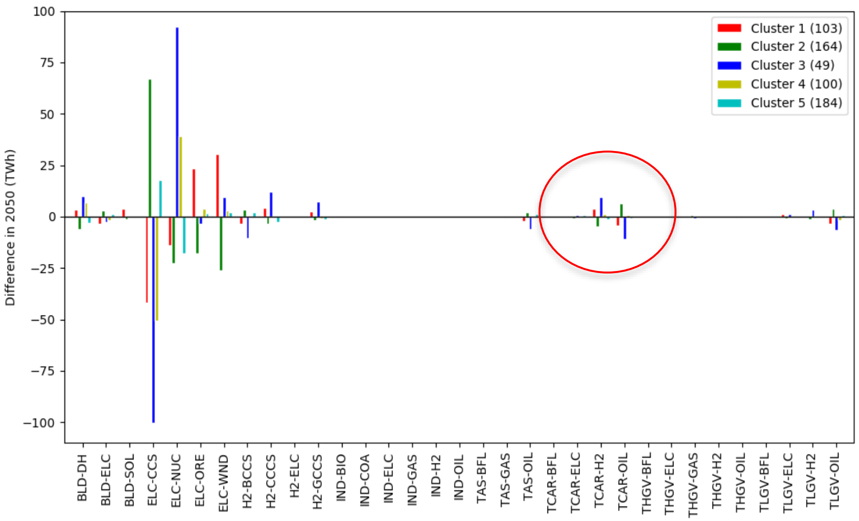
Sectoral standarisation



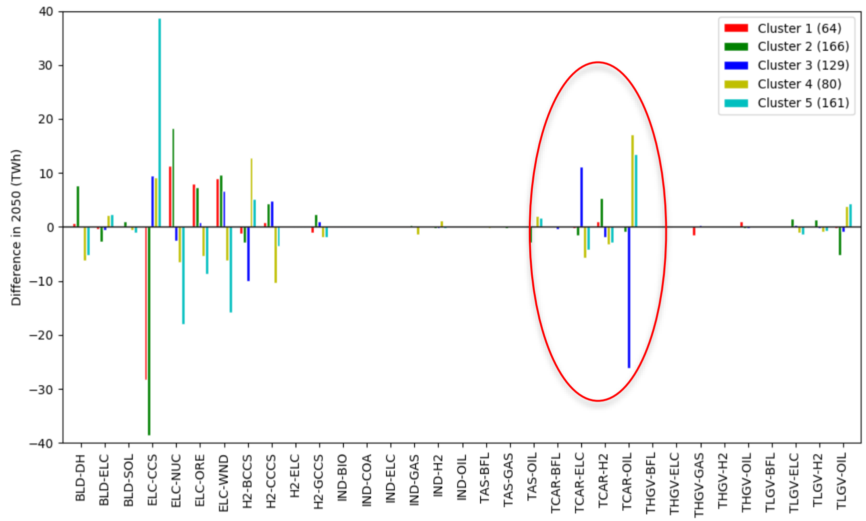
Treat every metric within a sector equally

Influence of Metric Transformation on Clustering

Metric transformation	Characteristics of clustered metrics
None	● All metrics are close to averages in 2030 ● Power sector is the only sector with obvious variation in 2050, as shown in Figure 6(a)
Power	● Influence of high variance of a few metrics, such as those in the power sector, is mitigated ● Variation of some metrics becomes more obvious, such as bioenergy consumption in the buildings (BLD-BIO) in 2030 and hydrogen production by gas and CCS (H2-GCCS) in 2050
Standardisation	● Variance of every metric is treated equally ● Variance of many clustered metrics is more obvious in both 2030 and 2050
Sectoral standardisation	● Relatively high variance of metrics in a sector are more likely to be revealed ● Trade-off between metrics in a sector is clearer, such as oil cars (TCAR-OIL) and EVs (TCAR-ELC) in 2050, as shown in Figure 6(b)



Original (2050)



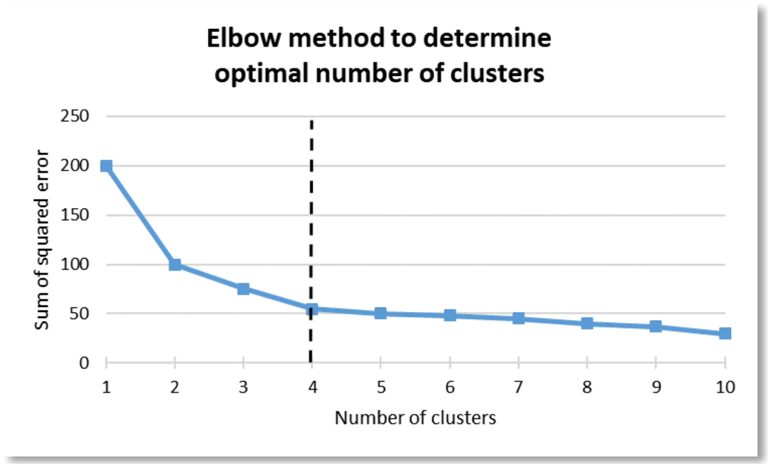
Sectoral standarisation (2050)

Performance evaluations and the choice of number of clusters

- No predefined categories for evaluation
- Criteria: **cohesion** and **separation**

Indicator	**Sum of squared error (SSE)**	**Davies-Bouldin (DB)**	**Calinski-Harabaz (CH)**	**Dunn (DN)**	**Silhouette (SL)**
Formula	$\sum_{c=1}^{C}\sum_{i\in c}\sum_{m=1}^{M}\lVert x_{i,m}$	$\frac{1}{C}\sum_{c=1}^{C}\max_{c'}\left\{\frac{S_c+S_{c'}}{\lVert\bar{x}_c-\bar{x}_{c'}\rVert}\right\}$	$\frac{\sum_{c=1}^{C} n_c\lVert\bar{x}_c-\bar{x}_g\rVert^2\ /\ (C}{\sum_{c=1}^{C}\sum_{i\in c}\lVert x_i-\bar{x}_c\rVert^2\ /\ (}$	$\frac{\min_{i\in c, j\in c'}\lVert x_i-x_j\rVert}{\max_{i,j\in c}\lVert x_i-x_j\rVert}$	$\frac{1}{N}\sum_{c=1}^{C}\sum_{i\in c}\frac{b_i-a_i}{\max\{b_i,a_i}$
Improvement direction	↓	↓	↑	↑	↑

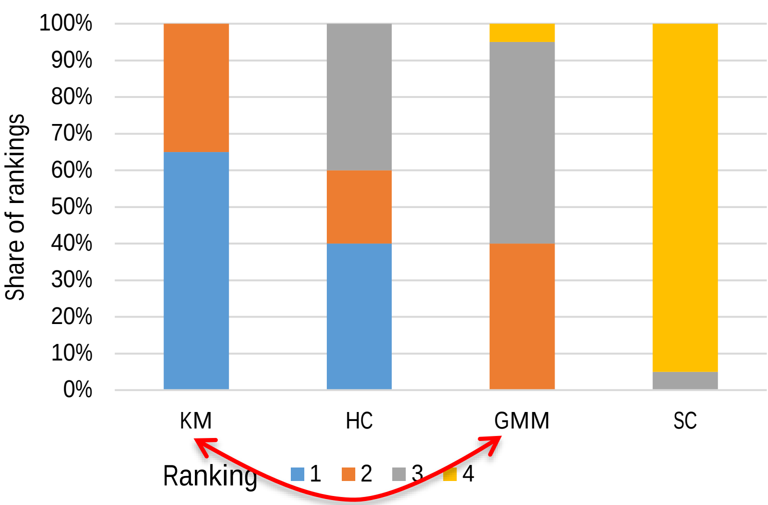
- Elbow measure
  - Turning point
  - Ideal number of clusters



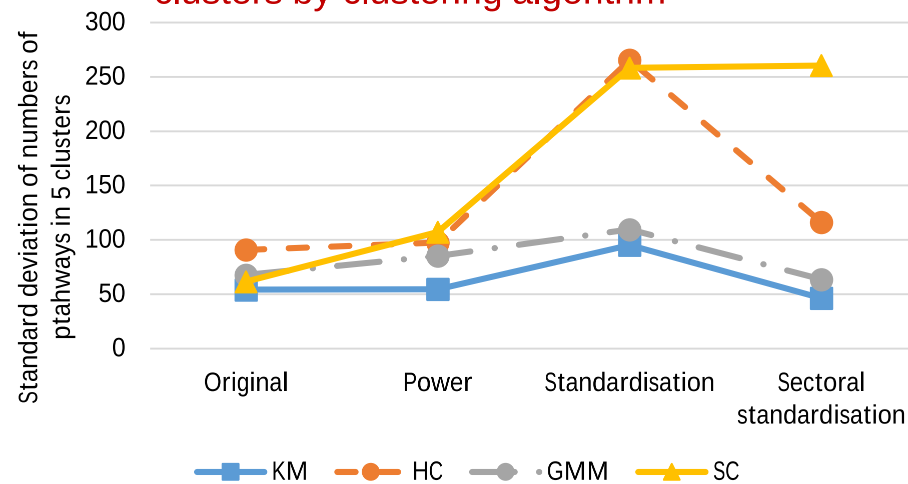
Performance comparison of clustering algorithms

Metric transformation	Clustering algorithm	Clustering validity index				
		SSE	DB	CH	DN	SL
None	KM	1	1	1	2	1
	HC	2	1	2	1	2
	GMM	3	3	3	3	3
	SC	4	4	4	4	4
Power	KM	1	1	1	2	1
	HC	3	2	3	1	3
	GMM	2	3	2	3	2
	SC	4	4	4	4	4
Standardisation	KM	1	2	1	2	2
	HC	3	1	3	1	1
	GMM	2	4	2	3	3
	SC	4	3	4	4	4
Sectoral standardisation	KM	1	2	1	2	1
	HC	3	1	3	1	3
	GMM	2	3	2	3	2
	SC	4	4	4	4	4

Performance of clustering algorithms

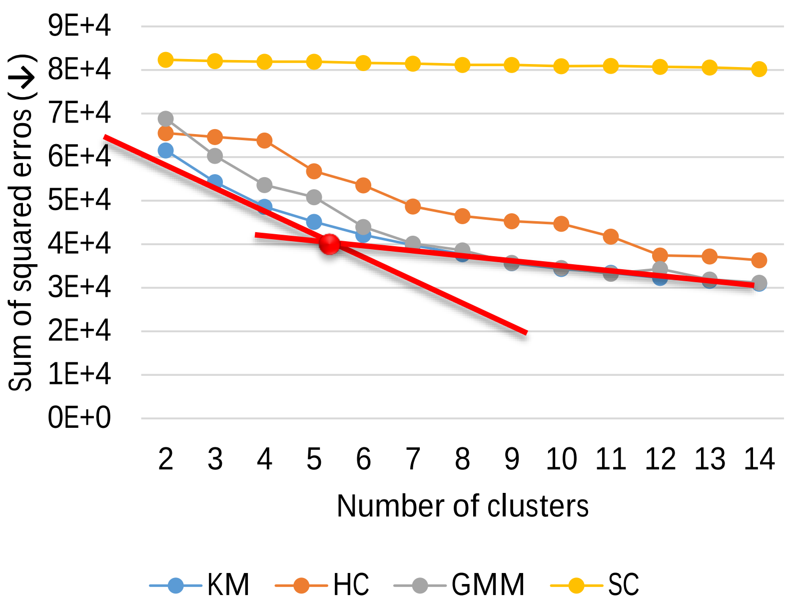


Distribution of pathways across identified clusters by clustering algorithm



Representative Long-term Decarbonisation Pathways

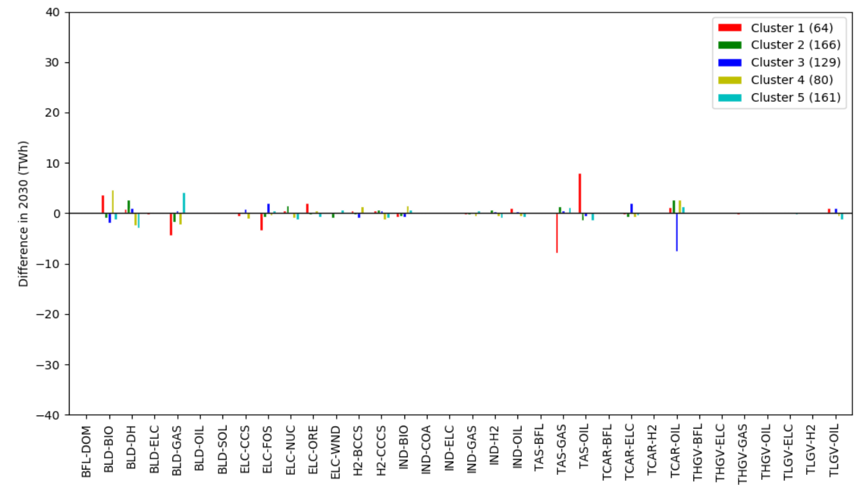
- **Elbow measure:**
- 5 clusters could be ideal for pathway characterisation



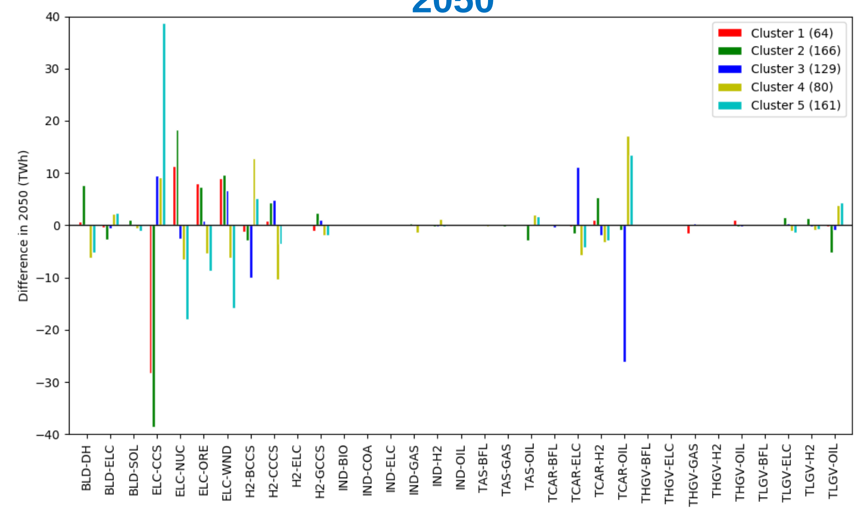
- Centroid pathways of clusters as representative pathways!

Average deviation of Metrics in representative pathways

2030



2050



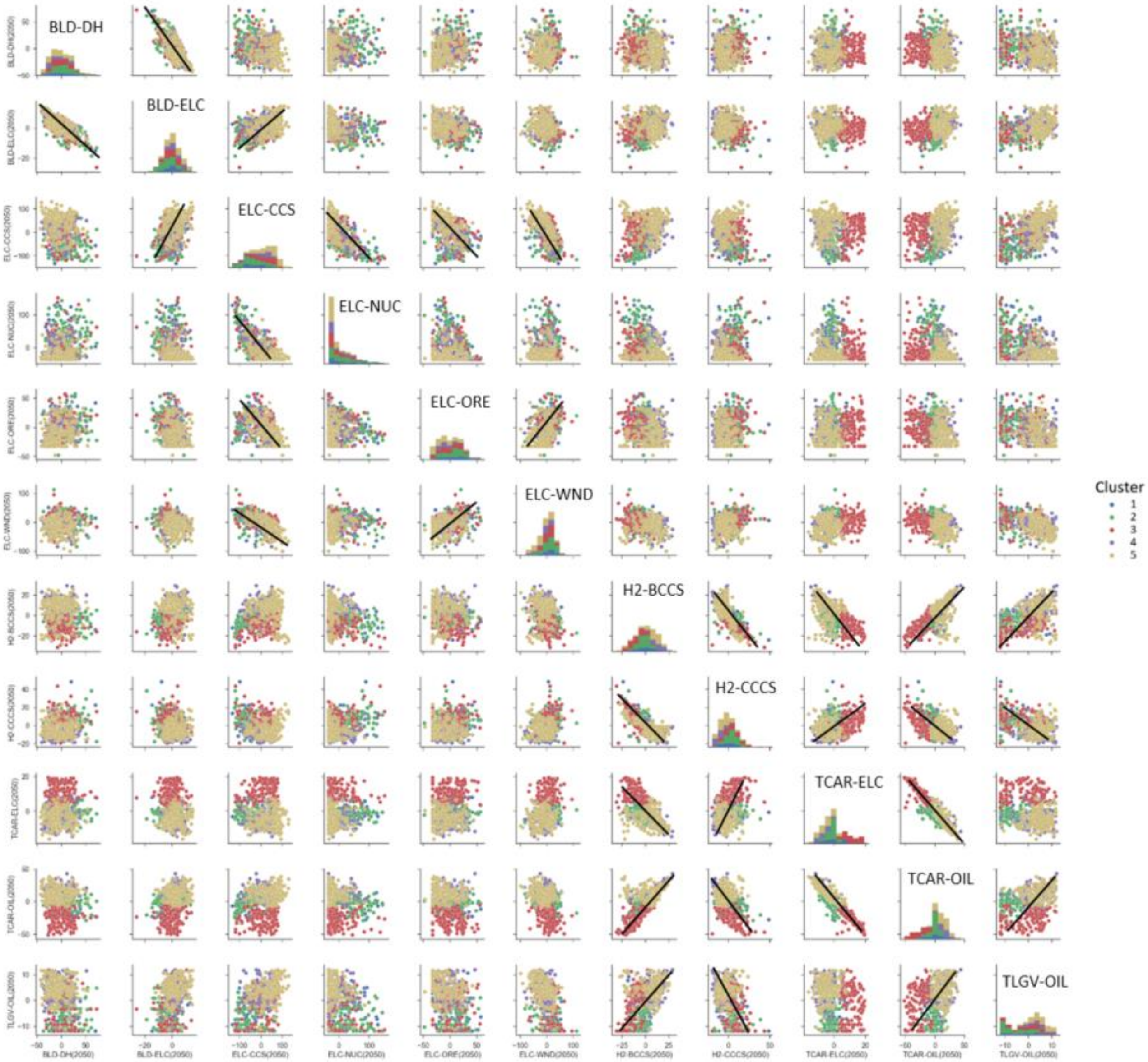
Identified Representative Pathways (2030 and 2050)

Key characteristics

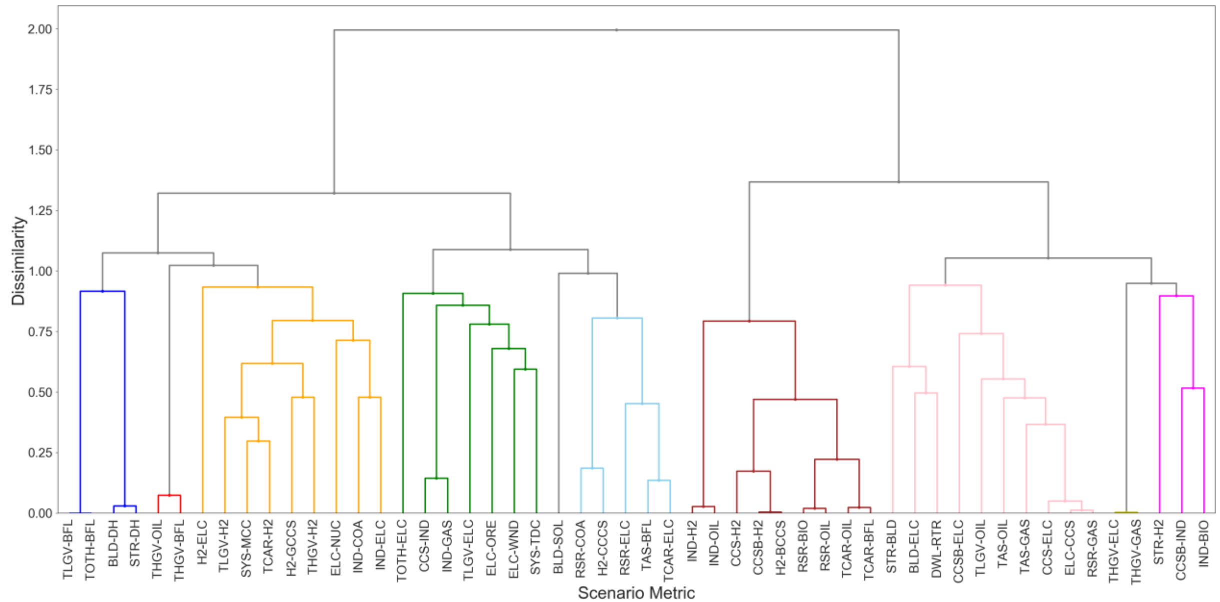
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Key characteristics	Lower CCS High low carbon elc	Higher district heating High low carbon elc Higher H2	Higher CCS Higher EV	Higher elc in building More CCS and BECCS More oil cars	Similar to #4 But more fossil fuel in 2030

Sector			Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
			2030	2050	2030	2050	2030	2050	2030	2050	2030	2050
Building		BIO	+				-		+		-	
		DH			+	++	+		-	--	-	--
		ELC				-				+		+
		GAS	-		-				-		+	
		SOL				+						
Electricity		CCS		---		---		++	-	++		+++
		FOS	-				+					
		NUC		+++	+	+++		-		--	-	---
		ORE	+	++		++				--		--
		WND		++		++		++		--		---
Hydrogen		BCCS		-		-		---	+	+++		++
		CCCS				+		+	-	---		-
		GCCS		-		+		+		-		-
Trans-port	AV	GAS	--		+						+	
		OIL	++		-	-				+	-	+
	CAR	ELC				-	+	+++		--		-
		H2		+		++		-		-		-
		OIL	+		+		--	---	+	+++	+	+++
	HGV	GAS		-								
		OIL		+								
	LGV	ELC				+				-		-
		H2				+						
		OIL	+			--	+			+	-	+

Correlation between key metrics in 2050



Coevolution of metrics across pathways with hierarchical clustering



Cluster colour	Cluster name	Cluster metrics	Negatively correlated clusters
Orange	H_2 production with gas for transport	H_2 production (via gas steam methane reforming (SMR)) and use in the transport sector.	Brown (-0.51)
Green	Renewable generation	Renewable power generation options, costs metrics, selected transport electrification.	Brown (-0.48)
Sky blue	Passenger car electrification	Passenger transport electrification; system electricity; aviation biofuels.	Brown (-0.66)
Brown	H_2 with bio CCS, car oil use	Biomass resource; H_2 production with CCS & bioenergy; oil in cars; system oil use; H2 and oil use in industry.	Orange (-0.51), Green (-0.48), Sky blue (-0.66)
Pink	Building electrification, power gen. w/ CCS	Electrification of buildings – as per the description in Table 2; CCS in power sector, and system gas use.	Blue (-0.94)
Blue	District heating	District heating (and storage). Clustered with transport biofuel use but weak correlation.	Pink (-0.94)

Pye, S, Li, P-H, Keppo, I, O'Gallachoir, B. 2019. Technology interdependency in the United Kingdom's low carbon energy transition, Energy Strategy Reviews, 24, 314-330.

Conclusions

- Performance of clustering algorithms is highly sensitive to the distribution of pathways
- For evenly distributed pathways (generated by Monte Carlo approach for uncertainty analysis), k-mean is the most robust choice
- Sectoral standarisation can emphasise key metrics in a sector without being overshadowed by a few key metrics with extremely high values in other sectors
- For coevolution of metrics across pathways, hierarchical clustering is useful to identify highly correlated metric sets
- More detailed technologies can be taken into account!
- Can be applied to characterise pathways from various models

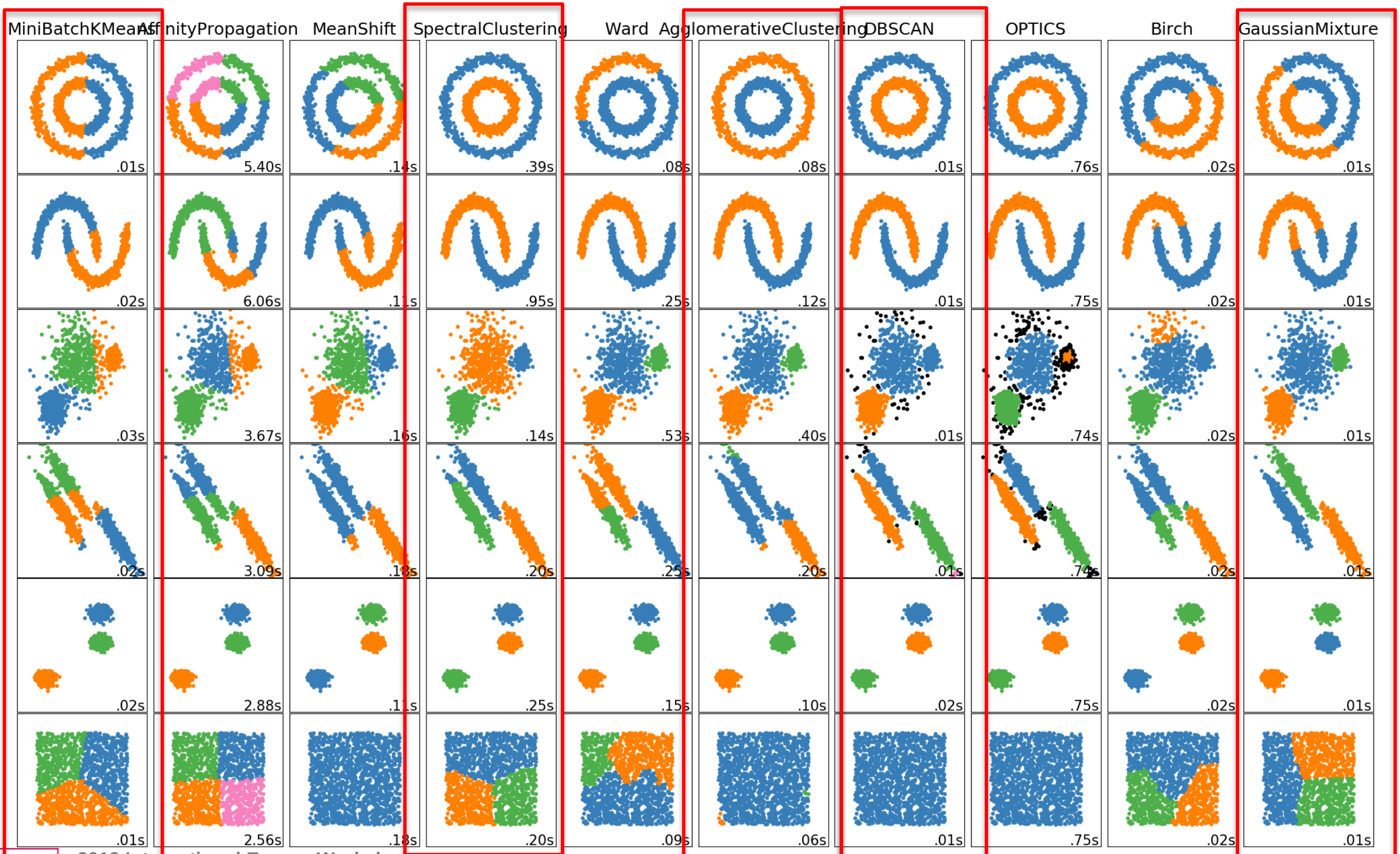
- Future tasks: characterise pathways with strong constraints, such as no CCS

Thanks for your attention!

Dr. Peihao Li (p.li@ucl.ac.uk)
Senior Research Fellow
UCL Energy Institute

Metric	Definition	Units
ELC-WND	Wind generation level	TWh
ELC-NUC	Nuclear generation level	TWh
ELC-CCS	CCS generation level	TWh
ELC-ORE	Other renewable generation level	TWh
ELC-FOS	Fossil generation level	TWh
BLD-BIO	Building bioenergy consumption	TWh
BLD-ELC	Building electricity consumption	TWh
BLD-GAS	Building gas consumption	TWh
BLD-OIL	Building oil consumption	TWh
BLD-DH	Building district heating consumption	TWh
BLD-SOL	Building solar energy consumption	TWh
H2-BCCS	H2 production by biomass gasification with CCS	TWh
H2-CCCS	H2 production by coal gasification with CCS	TWh
H2-ELC	H2 production by electrolysis	TWh
H2-GCCS	H2 production by gas (SMR) with CCS	TWh
H2-GAS	H2 production by gas (SMR)	TWh
IND-BIO	Industry bioenergy consumption	TWh
IND-COA	Industry coal consumption	TWh
IND-ELC	Industry electricity consumption	TWh
IND-GAS	Industry gas consumption	TWh
IND-H2	Industry hydrogen consumption	TWh
IND-OIL	Industry oil consumption	TWh
TAS-GAS	Aviation & shipping - gas	TWh
TAS-OIL	Aviation & shipping - oil	TWh
TAS-BFL	Aviation & shipping - biofuel	TWh
TCAR-ELC	Cars - electricity	TWh
TCAR-GAS	Cars - gas	TWh
TCAR-H2	Cars - H2	TWh
TCAR-OIL	Cars - oil	TWh
TCAR-BFL	Cars - biofuels	TWh
THGV-ELC	Heavy goods vehicles - electricity	TWh
THGV-GAS	Heavy goods vehicles - gas	TWh
THGV-H2	Heavy goods vehicles - H2	TWh
THGV-OIL	Heavy goods vehicles – oil	TWh
THGV-BFL	Heavy goods vehicles - biofuels	TWh
TLGV-ELC	Light goods vehicles - electricity	TWh
TLGV-H2	Light goods vehicles - H2	TWh
TLGV-OIL	Light goods vehicles - oil	TWh

Comparison between clustering algorithms



Source: scikit-documentation: 2.3 clustering https://scikit-learn.org/stable/modules/clustering.html